

Die Rolle der Wahrscheinlichkeit für das Verständnis beurteilender Statistik

Manfred Borovcnik

Universität Klagenfurt, Institut für Statistik

Zusammenfassung

Die Wahrscheinlichkeit ist die Grundlage für intelligente Handlungen und Entscheidungen angesichts von Ungewissheit. Dazu gehören statistische Schlussfolgerungen ebenso wie Überlegungen zu Zuverlässigkeit, Risiko und Entscheidungsfindung.

Die Lehrpläne haben den Zugang zur Wahrscheinlichkeitsrechnung und damit die Natur von Wahrscheinlichkeit verengt.

Mit dem Zugang zu Technologie ist die Simulation zum vorherrschenden Lehransatz geworden.

Obwohl Simulation eine wirksame Methode ist, um komplizierte Mathematik zu ersetzen, reduziert sie die Konzepte auf ihren frequentistischen Anteil.

Dies gipfelt in einem Ansatz zur Informellen Inferenz, der Wahrscheinlichkeit und bedingte Wahrscheinlichkeit überflüssig macht, um statistische Beurteilung zu unterrichten.

Die relevanten Eigenschaften statistischer Inferenz erfordern jedoch, dass sich im kognitiven System des Einzelnen eine universelle Vorstellung von Wahrscheinlichkeit herausbildet.

Überblick

1. Fünf Säulen für die Wahrscheinlichkeitsrechnung und damit für die Statistik

1. Beginne mit Wahrscheinlichkeit so früh wie möglich

2. Nutze Spiele auf intelligente Weise

3. Forme Bayesianisches und risiko-orientiertes Denken

4. Verknüpfe Wahrscheinlichkeitsrechnung und beurteilende Statistik

5. Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko

2. Statistische Inferenz verstehen lernen mittels bedingter Wahrscheinlichkeit

- Bedingte Wahrscheinlichkeit ist der Schlüssel für statistische Inferenz
- Die Fisher-Neyman-Pearson-Kontroverse
- Interpretation von Fehlentscheidungen bei statistischen Tests auf lange Sicht
- Analogie zwischen Medizin und statistischen Tests

Einleitung

Fünf Säulen für die Unterweisung in Wahrscheinlichkeit

- **Um die Bemühungen in Didaktik und Unterricht zu strukturieren**
 - Damit der vielfältige Begriff Wahrscheinlichkeit und sein Zweck klarer werden.
- **Die Schwierigkeiten im Zugang zur beurteilenden Statistik zu lösen**
 - Die verschiedenen Deutungen von Wahrscheinlichkeit verbinden,
 - Wahrscheinlichkeit und statistischer Beurteilung miteinander verknüpfen,
 - und nachhaltige Intuitionen aufbauen, die helfen, Gelerntes zu verstehen und zu behalten.

1. Fünf Säulen für die Wahrscheinlichkeitsrechnung und damit für die Statistik

1. Beginne mit Wahrscheinlichkeit so früh wie möglich

und entwickle Ideen und Konzepte spiralförmig.

2. Nutze Spiele auf intelligente Weise,

um nachhaltige probabilistische Intuitionen aufzubauen.

3. Forme Bayesianisches und risiko-orientiertes Denken

so früh wie möglich.

4. Verknüpfe Wahrscheinlichkeitsrechnung und beurteilende Statistik

von der ersten Einführung in Wahrscheinlichkeit an.

5. Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko

im Sinne von Zwillingbegriffen.

Baue den Begriff der bedingten Wahrscheinlichkeit sorgfältig auf und ermögliche eine umfassendere Konzeption statistischer Beurteilung.

Wahrscheinlichkeit und statistische Beurteilung verknüpfen: Zweck und Begründung

- Auf der reichen Erfahrung aus Spielen aus Kindheit und Jugend aufbauen, um zuverlässige Intuitionen zu entwickeln (#1, #2).
- Die subjektivistische (epistemische) Konnotation von Wahrscheinlichkeit miteinbeziehen, um spätere Verwirrung zu vermeiden und die Anwendungen zu erweitern (#3).
- Mit Fragen der beurteilenden Statistik in den frühesten Phasen des Unterrichts in Wahrscheinlichkeit verknüpfen, weil es sich um komplementäre Teile desselben handelt (#4).
- Wahrscheinlichkeit und Risiko als Zwillingsbegriffe entwickeln (#5).

Die Kontroverse in den Grundlagen (sind klassische oder Bayesianische Methoden besser)

➤ für einen pluralistischen Hintergrund nützen statt Methoden zu reduzieren.

Den Blickwinkel auf Intuitionen und Common Sense schärfen.

➤ Bedingte Wahrscheinlichkeit ist der Schlüssel zu einer breiteren Interpretation der Inferenz.

Probabilistische Literalität – Gewandtheit in probabilistischen Fragen

- Verständnisschwierigkeiten sind notorisch
 - ✓ bei statistischen Schlussfolgerungen und der Wahrscheinlichkeitsrechnung.

Eine beliebte Intervention im Unterricht besteht darin

- ✓ zu vereinfachen – siehe aber Brousseaus „glissement didactique“.

Innovative Reaktion:

- ✓ Wahrscheinlichkeit mit Fragen statistischer Inferenz verknüpfen.
- ✓ Die Konzeption der Begriffe erweitern statt den theoretischen Rahmen einschränken.

Wahrscheinlichkeit ist viel mehr als nur relative Häufigkeiten und hat eher einen metaphorischen als materiellen Charakter.

1. Beginne mit Wahrscheinlichkeit so früh wie möglich und entwickle Ideen und Konzepte spiralförmig

Kreative Aufgaben, die den Zweck von Wahrscheinlichkeit sichtbar machen (Varga, 1983).

Varga untersuchte mit 9-Jährigen das Verhalten des Zufalls in Bezug auf Runs von Kopf und Zahl bei Münzwürfen.

- ✓ Nicht um zu untersuchen, wie sich relative Häufigkeiten und singuläre Muster entwickeln.
- ✓ Sondern um zu beurteilen, ob ein bestimmtes Münzwurfprotokoll echt oder „erfunden“ war.

Diese Aufgabe stellt die Kinder schon zu Beginn des Unterrichts in Wahrscheinlichkeit in den Mittelpunkt von Überlegungen zur statistischen Beurteilung.

1. Frühe Ideen und Konzepte

um das Treffen von Entscheidungen mit statistischer Beurteilung zu verbinden

Es gilt, die wichtigsten Zwecke von Wahrscheinlichkeit sichtbar zu machen:

- ✓ Entscheidungen zu treffen & eine statistische Beurteilung von Möglichkeiten abzugeben.

In beiden Situationen geht es um Risiko und Formen, mit Unsicherheit umzugehen.

- Es geht darum, Wahrscheinlichkeit in Spielen sichtbar und relevant werden zu lassen.
- Den Kindern Zeit zu geben, dass in ihnen Begrifflichkeiten wachsen und sie ihre Vorstellungen abändern und den Gegebenheiten anpassen können.

“Vertrautheit mit kombinatorischem Denken lässt ... Hypothesentests ... machbar werden. ... Man kann Testen von Hypothesen darauf aufbauen. ... **es ist didaktisch sinnvoll, ... statistische Tests ... über Kombinatorik ... zu erschließen**”

(Fejes-Tóth, 2022).

#2. Nutze Spiele auf intelligente Weise, um nachhaltige probabilistische Intuitionen aufzubauen

Man setze Glücksspiele nicht routinemäßig, sondern auf intelligente Weise ein, so wie es Varga (1983) getan hat.

Spiele sind nützlich, um Verbindungen zwischen den zentralen Bedeutungen von Wahrscheinlichkeit herzustellen:

- ✓ die klassische Interpretation von Proportionen
als Gewichtung der Tendenz, Ergebnisse zu produzieren – ex ante;
- ✓ die frequentistische Bedeutung von Wahrscheinlichkeit
als ein Maß für den Zufall – ex post;
- ✓ Verhältnisse von Chancen und Einsätzen
zur Kalibrierung der subjektivistischen (epistemischen) Wahrscheinlichkeit.

Alle bedingen ihre je eigenen Voraussetzungen.

#2. Nutze Spiele auf intelligente Weise, um Deutungen von Wahrscheinlichkeit durch Aufgaben unter inferenziellem Blickwinkel miteinander zu verbinden

Die Interpretationen von Wahrscheinlichkeit haben je ihre eigenen Voraussetzungen.

- Gleichwahrscheinliche Elementarereignisse.
- Unabhängig voneinander unter gleichen Bedingungen wiederholbare Experimente .
- Ein rationales System von Präferenzen für eine Person.

Steinbring (1991) verweist

- ✓ auf eine Komplementarität zwischen den Konzepten der Gleichwahrscheinlichkeit und der Wahrscheinlichkeit als etwas wie eine relative Häufigkeit,
- ✓ darauf, dass die Konzeption der Wahrscheinlichkeit **beide Aspekte** erfordert.

Die **Wechselwirkung** zwischen der auf gleichwahrscheinlichen Fällen beruhenden Wahrscheinlichkeit und der Entwicklung relativer Häufigkeiten **führt zu Fragen der statistischen Inferenz.**

#2. Nutze Spiele auf intelligente Weise, um probabilistische Intuitionen zu erweitern

Interessanterweise wird bei statistischer Inferenz in der Regel von der Unabhängigkeit der einzelnen Spiele ausgegangen und es werden keinerlei Muster untersucht.

- Die Untersuchung solcher Muster wäre ja sinnlos, da der Zufall bedeutet, dass alles passieren kann.

Es handelt sich um eine wechselseitige Abhängigkeit:

- Fragen der statistischen Inferenz sind Schlüssel zu einem guten Verständnis der Wahrscheinlichkeitsrechnung.
- Ein angemessenes Konzept der Wahrscheinlichkeit ist der Schlüssel zum Verständnis statistischer Schlussfolgerungen.

Das Spiel integriert relative Risiken und die Messung des Grads des Vertrauens:

Der Übergang vom komplementären Paar – Laplace-Wahrscheinlichkeiten und relative Häufigkeiten – zum unauflösbaren Tripel.

3. Forme Bayesianisches und risiko-orientiertes Denken so früh wie möglich

Bayesianische Ideen beziehen sich auf

- ✓ intuitives Denken und Common Sense;
- ✓ bedingte Wahrscheinlichkeit, welche die Folgen (Kosten) eines ungewissen Ereignisses mit einbezieht.

Carranza und Kuzniak (2008)

- weisen auf die problematische Natur der bedingten Wahrscheinlichkeit hin;
- sie weisen auf Probleme hin, welche auf die Vernachlässigung der subjektivistischen Auslegung von Wahrscheinlichkeit zurückzuführen sind.

Ein qualitatives Urteil über eine Wahrscheinlichkeit basiert auf dem Präferenzsystem einer Person und ist daher ein mathematischer Ausdruck dieser Präferenzen.

3. Forme Bayesianisches und risiko-orientiertes Denken: subjektivistisch = epistemisch ≠ beliebig

Ein qualitatives Urteil über die Wahrscheinlichkeit einer Aussage wird oft dahingehend missverstanden, dass es sich um einen willkürlichen Wahrscheinlichkeitswert handelt.

Der wesentliche „Unterschied“ zu anderen Bedeutungen von Wahrscheinlichkeit besteht darin, dass in der gängigen Vorstellung

- das Urteil einer Person subjektiv (oder sogar willkürlich) sein „muss“,
- die Eigenschaft eines Prozesses oder Geräts dagegen „objektiv“ (wissenschaftlich und unbestreitbar) wäre.

Das Urteil eines Menschen muss jedoch auf qualitativen Kenntnissen beruhen und ist daher keineswegs willkürlich.

3. Forme Bayesianisches und risiko-orientiertes Denken: Die Diskussion rund um Bayes im Unterricht

Diskussion über den Einsatz von Bayes-Methoden im Unterricht (*American Statistician*, Teachers' Corner, 1997):

- ✓ Verschiedene Interpretationen von Wahrscheinlichkeit und ihre Rechtfertigung durch eine axiomatische Theorie.
- ✓ Die Bayes-Formel ist der Schlüssel zu jedem statistischen Verfahren der Inferenz.
- ✓ Das Verständnis statistischer Schlussfolgerungen erfordert eine gute Kenntnis der bedingten Wahrscheinlichkeit und ein ausgewogenes Konzept der Wahrscheinlichkeit.

Dies schließt sowohl die Gleichwahrscheinlichkeit als auch die frequentistische und die subjektivistische Auslegung von Wahrscheinlichkeit ein.

3. Forme Bayesianisches und risiko-orientiertes Denken, um beurteilende Statistik zu verstehen

- Bayes-Methoden sind nützlich, um die berüchtigten Fehlinterpretationen der statistischen Inferenz zu vermeiden:

„Der Student ... kann das Signifikanzniveau nur deshalb nach Bayes falsch interpretieren, weil er nichts über eine Bayesianische Alternative zum Signifikanztest gelernt hat“ (Diepgen, 1992).

- Bayesianische Probleme verknüpfen auf natürliche Weise Wahrscheinlichkeit mit Risiko.

Das Risiko überlappt mit persönlichen Überzeugungen über Methoden und dem jeweiligen Kontext bzw. der individuellen Auffassung darüber:

eine Herausforderung!

4. Verknüpfung Wahrscheinlichkeitsrechnung und beurteilende Statistik von der ersten Einführung in Wahrscheinlichkeit an

Eine Schätzung der unbekanntes Wahrscheinlichkeit ist erforderlich

- für Spiele mit unbekannter Struktur
- für allgemeine Zufallsprozesse (die über Glücksspiele hinausgehen).

Eine solche Schätzung erfordert eine frequentistische Interpretation der Wahrscheinlichkeit.

Um die Beziehung zwischen Wahrscheinlichkeit und relativen Häufigkeiten zu klären, wird schon im Anfangsunterricht von Wahrscheinlichkeit viel Wert auf ein geeignetes Verständnis des **empirischen Gesetzes der großen Zahlen** gelegt.

Allerdings mit ungeeignetem Blickwinkel auf die Konvergenz der relativen Häufigkeiten.

Vielmehr sollte man den Zusammenhang mit Fragen der statistischen Inferenz ausbreiten und nutzen.

4. Verknüpfte Wahrscheinlichkeitsrechnung und beurteilende Statistik weit über Spiele hinaus

Die Schätzung oder Beurteilung von Hypothesen für unbekannte Wahrscheinlichkeiten wirft folgende Fragen auf:

- Welcher Wahrscheinlichkeitswert ist für ein untersuchtes Ereignis gerechtfertigt?
- Wie kann man behaupten, über genügend Daten zu verfügen, damit eine empirische Wahrscheinlichkeitsschätzung gut genug ist?

Diese Fragen bringen die Wahrscheinlichkeit notwendigerweise mit statistischer Inferenz in Verbindung.

Auf diese Weise wird deutlich, dass die statistische Inferenz den Fokus auf weitere Interpretationen der Wahrscheinlichkeit lenkt, wenn man diese Fragen wirklich klären will.

4. Verknüpfung Wahrscheinlichkeitsrechnung und beurteilende Statistik : Inmitten der Bayes-Kontroverse über die Grundlagen von Wahrscheinlichkeit

Für die statistische Inferenz ist es zu wenig, von der Gleichwahrscheinlichkeit zu einer frequentistischen Konzeption von Wahrscheinlichkeit überzugehen:

- Um entscheidende Fehler zu vermeiden, musste man eine subjektivistische Konnotation akzeptieren – oder diese Fehler billigend in Kauf nehmen (Hacking, 1965).

Die Kontroverse in den Grundlagen der Wahrscheinlichkeitsrechnung (1930-80er Jahre)

- spiegelt sich in dem Dilemma wider, das Carranza und Kuzniak (2008) für den Unterricht feststellen.

Neben der Debatte darüber, welche Interpretation von Wahrscheinlichkeit die bessere ist, wurde es dringend notwendig, die Komplexität von statistischer Inferenz zu reduzieren, um sie unterrichten zu können.

4. Verknüpfte Wahrscheinlichkeitsrechnung und beurteilende Statistik: Reduktion der Komplexität

Cobb (2007) schlug vor, die statistische Inferenz vollständig durch Resampling zu ersetzen. So wurde ein Ansatz entwickelt, der als informelle (simulationsbasierte) Inferenz bezeichnet wird.

- Diese so genannte **Informelle Inferenz** reduziert den Begriff der Wahrscheinlichkeit auf seinen frequentistischen Aspekt und macht Wahrscheinlichkeit eigentlich überflüssig, da alles durch Simulation gelöst wird, indem die Daten gemischt und dann einige zufällig ausgewählt werden, wobei dieser Prozess viele Male wiederholt wird.
- Batanero und Borovcnik (2016) dagegen konzentrieren sich auf **Szenarien, die in einen Kontext eingebettet sind**, der die Komplexität auf natürliche Weise reduziert und eine intuitiv zugängliche Bedeutung für die beteiligten Konzepte hat. Natürlich kann zur Veranschaulichung der Eigenschaften der Konzepte auch Simulation verwendet werden.

4. Verknüpfte Wahrscheinlichkeitsrechnung und beurteilende Statistik: Reduktion der Komplexität

Die Auffassung von Güteindizes für statistische Tests als bedingte Wahrscheinlichkeiten

- ✓ wird durch den Szenario-Ansatz von Batanero und Borovcnik (2016) untermauert;
- ✓ wird dagegen im Rahmen der informellen Inferenz zu absoluten (also nicht-bedingten) Wahrscheinlichkeiten degeneriert.

Eine Kritik an der informellen Inferenz – nicht alle Indizes/Fehler können in diesem Ansatz berücksichtigt werden – findet man in Borovcnik (2021).

Der Szenario-Ansatz extrahiert eine natürliche Interpretation von Konzepten aus dem Kontext.

Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko im Sinne von Zwillingsbegriffen

Wahrscheinlichkeit und Risiko haben sich im Gleichschritt auf ähnliche Weise entwickelt, was ihre Abgrenzung voneinander erschwert.

In der Literatur zum Thema Risiko werden uneinheitliche Begriffe verwendet (Borovcnik und Kapadia, 2018).

Die Frage ist, was Risiko denn eigentlich umfassen sollte:

- die Wahrscheinlichkeit eines „unerwünschten“ Ereignisses **und** dessen Auswirkungen (Kosten, Nutzen),
- oder **nur die Wahrscheinlichkeit** des unerwünschten Ergebnisses
- oder **nur die Auswirkungen** (Kosten, Nutzen),
- oder ob sich das Risiko **indirekt auf die Faktoren im Hintergrund – Hazards genannt - bezieht**, die potenziell das unerwünschte Ergebnis „verursachen“.

Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko

Der Fall der kleinen Wahrscheinlichkeiten

Wenn Wahrscheinlichkeiten oder Risiken sehr gering, die Auswirkungen aber sehr groß sind, neigen Menschen dazu, die geringe Wahrscheinlichkeit zu ignorieren und ihre Entscheidungen ausschließlich auf den potenziellen Nutzen oder Schaden zu stützen.

- Aus diesem Grund spielen Menschen im Lotto.

Sie glauben an das Urteil Gottes oder erwarten Gottes Unterstützung, um im Lotto zu gewinnen.

- Deshalb schließen Menschen fast jede Versicherung ab.

Sie scheinen zu glauben, dass sie durch Abschluss einer Versicherung über den finanziellen Verlust hinaus einen persönlichen Schaden vermeiden können.

Oder sie „versuchen, Gott zu besänftigen“, indem sie die Versicherung abschließen.

Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko

Kleine Wahrscheinlichkeiten und die moralische Wahrscheinlichkeit

In Risikobetrachtungen tauchen sehr kleine Wahrscheinlichkeiten auf; diese sind Modellgrößen, die anderweitig (durch Modellannahmen) begründet werden müssen.

Kleine Wahrscheinlichkeiten können selbst unter reinen Laborbedingungen (Simulation) nicht richtig geschätzt werden. Noch viel weniger können vernünftige Schätzungen in realen Situationen vorgenommen werden.

Daher wurde in der Vergangenheit **das Konzept der moralischen Wahrscheinlichkeit in Betracht gezogen, eine Schranke, unterhalb derer alle Wahrscheinlichkeiten auf Null zu setzen sind.**

Als Schwellenwert wurden Werte in der Größenordnung von 10^{-4} diskutiert.

Moderne Anwendungen mit Risiken beginnen erst weit unterhalb dieser Schwelle und sind im Sinne der moralischen Wahrscheinlichkeit merkwürdig.

Die Schätzung einer kleinen, unbekanntes Wahrscheinlichkeit bleibt auch bei enorm vielen Daten unpräzise und ungewiss

Aufgabe ist es, eine unbekanntes Wahrscheinlichkeit von $p = 0.0001$ mittels 10 000 Daten zu schätzen

Ergebnis von 100 Schätzungen	Schätzung von p	Relativer Fehler der Schätzung	Anzahl	%	Wahrscheinlichkeit	
	0	0.0000	-100%	41	41.0%	0.3679
	1	0.0001	0%	34	34.0%	0.3679
	2	0.0002	100%	20	20.0%	0.1839
	3	0.0003	200%	3	3.0%	0.0613
	4	0.0004	300%	2	2.0%	0.0153
	5	0.0005	400%	0	0.0%	0.0031
	6	0.0006	500%	0	0.0%	0.0005
	7	0.0007	600%	0	0.0%	7.29E-05
	8	0.0008	700%	0	0.0%	9.11E-06
	9	0.0009	800%	0	0.0%	1.01E-06
	10	0.0010	900%	0	0.0%	1.01E-07
				0		
				100	100.0%	1.0000

Das ist der **wahre Wert von p** , den man aus 10 000 Daten schätzen soll.

Entwickle die enge Verwandtschaft zwischen Wahrscheinlichkeit und Risiko: Psychologische Fragen, die sich aus der Ähnlichkeit ergeben

Folgender Aspekt von Risiko verwischt die Wahrnehmung von Wahrscheinlichkeiten :

- Es ist schwierig, wenn nicht gar unmöglich, die Bewertung einer Wahrscheinlichkeit von den Auswirkungen zu trennen.
- Wenn die Auswirkungen groß sind, neigen Menschen dazu, die Wahrscheinlichkeit des Ereignisses zu vernachlässigen.

Es gibt einen Unterschied im Verhalten der Menschen; nach Kahneman und Tversky (1979)

- sind Menschen in Verlustsituationen risikofreudig;
- dagegen sind sie in Gewinnsituationen risikoscheu.

Diese psychologischen Verzerrungen zeigen, wie schwierig statistische Inferenz ist, abgesehen von den mathematischen und methodischen Problemen.

Es ist unmöglich, Nutzen und Wahrscheinlichkeiten voneinander zu trennen

Kahneman & Tversky (KT) machten Experimente mit Erwachsenen. Sie interpretierten deren Antworten, als ob sie irrige Strategien verfolgten: Repräsentativität, Verfügbarkeit, Ankern, etc. **KT zeigten die Inkohärenz des Verhaltens von Individuen in Situationen mit Unsicherheit.**

	Gewinnsituation		Verlustsituation	
Gewinn [in \$]	Option a₁	Option a₂	Option a₃	Option a₄
Zukunft	1000	2500 mit ½ 0 mit ½	-1000	-2500 mit ½ 0 mit ½
	Viele ziehen die sicheren 1000 vor, auch wenn a ₂ einen Erwartungswert von 1250 bietet.		Viele ziehen a₄ vor; sie suchen das Risiko anstatt beim sicheren Verlust von -1000 zu bleiben.	

Charakterisierung von KT:

Risikoaversion in Gewinnsituationen und Risikoherausforderung in Verlustsituationen im Vergleich zum rationalen Verhalten gemäß dem Erwartungswert des Gewinns.

Es ist aber eine Verlustsituation – die Leute entscheiden intelligenter als KT!

MB	„Verlustsituation“	Gewinnsituation	KT
Nichts tun	Option a_2^*	Option a_1	Option a_2
1000 gehören schon mir!	1500 mit $\frac{1}{2}$ –1000 mit $\frac{1}{2}$	1000	2500 mit $\frac{1}{2}$ 0 mit $\frac{1}{2}$

KT: Wert (a_2) > Wert (a_1), daher ist a_2 optimal. Aber die Leute wählen meist a_1 ; sie sind daher **“in Gewinnsituationen fälschlicherweise risikoscheu”**.

MB: Man hat die 1000 schon sicher. Der zusätzliche Gewinn beträgt nur 1500, birgt aber das Risiko, die 1000 zu verlieren und bei 0 zu landen.

Es lohnt die Mühe nicht, die 1000 zu riskieren, der zusätzliche Gewinn ist zu klein. Man hat eigentlich eine **Verlustsituation**. Man muss einer Person, die schon 1000 hat, mehr anbieten, damit sie das Risiko auf sich nimmt, alles zu verlieren.

Wenn eine der Optionen keinen Zufall birgt, neigen Leute dazu, **die Optionen mit der Ausgangssituation zu vergleichen**. Das führt zu anderen Entscheidungen.

Die Leute sind intelligenter und keineswegs irrational, wie das KT einschätzen!

„Gewinnsituation“	MB	KT	Verlustsituation
Nichts tun	Option a ₄ *	Option a ₃	Option a ₄
-1000 gehören schon mir!	-1500 mit ½ 1000 mit ½	-1000	-2500 mit ½ 0 mit ½

KT: Nun, Wert (a₃) > Wert (a₄), deswegen ist a₃ optimal. Die Leute wählen meist a₄; “sie suchen in Verlustsituationen fälschlicherweise das Risiko”.

MB: Man hat ja schon einen sicheren Verlust von -1000. Jedoch bietet sich eine Möglichkeit an, die Schulden gänzlich loszuwerden, wenngleich damit das Risiko eines zusätzlichen Verlusts von 1500 verbunden ist! Eine Person, die schon 1000 an Schulden hat, wird jede Gelegenheit suchen, auf 0 zu kommen, und daher auch das Risiko eines weiteren Verlusts in Kauf nehmen. **Es ist eher eine Gewinnsituation.**

Diese Interpretation wird durch dieselbe Studie von KT bestätigt.

**In wiederholten Situationen existiert praktisch kein Risiko,
in Einzelentscheidungen sind Individuen dem totalen Risiko ausgesetzt**

Experiment 1^w: Gewinn [\$]

Gewinnsituation

Experiment 2^w: Gewinn [\$]

Verlustsituation

Option b ₁	Option b ₂
1 mit ½	2.5 mit ½ 0 mit ½
Option b ₃	Option b ₄
-1 mit ½	-2.5 mit ½ 0 mit ½

- **Einzelentscheidung** (alles auf einmal, multipliziert mit 1000, d.h. wie zuvor): **KT (MB)** in **Gewinnsituationen**: **b₂ (b₁)**. In **Verlustsituationen**: **b₃ (b₄)**
- **1000 einzelne Entscheidungen**: **Gewinnsituation**: Um mehr als 1000 zu gewinnen, muss man öfter als 400 Mal mit **b₂** gewinnen. Wahrscheinlichkeit: $1-10^{-10}$. **Verlustsituation**: Optimal ist nun **b₃**, mit **b₄** verliert man “sicher” 1000.

Bei wiederholten Entscheidungen spielt Nutzen gar keine Rolle. **Der KT-Blickwinkel passt nur für wiederholte Entscheidungen, versagt aber für Einzelentscheidungen!**

2. Statistische Inferenz verstehen lernen mittels bedingter Wahrscheinlichkeit

- **Der Schlüssel, statistische Inferenz verstehen zu lernen,**
ist, den Begriff der bedingten Wahrscheinlichkeit korrekt zu verstehen
- **Die Fisher-Neyman-Pearson-Kontroverse**
Die Debatte über eine frequentistische oder nicht-frequentistische Interpretation von Qualitätsindizes – die eigentlich bedingte Wahrscheinlichkeiten sind – von statistischen Tests.
- **Interpretation von Fehlentscheidungen bei statistischen Tests auf lange Sicht**
Der springende Punkt ist eine langfristige Interpretation von bedingten Wahrscheinlichkeiten, während sie eigentlich im Einzelfall verwendet werden.
- **Analogie zwischen Medizin und statistischen Tests**
Auch hier handelt es sich um eine Wechselwirkung von bedingten Wahrscheinlichkeiten mit dem Kontext: Der medizinische Kontext hilft, Begriffe und die Entscheidungssituation zu klären.

Statistische Inferenz verstehen lernen mittels der bedingten Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit spielt sowohl bei Fishers Signifikanztest als auch bei der Neyman-Pearson-Testpolitik eine grundlegende Rolle.

In Analogie dazu sind die „inversen“ bedingten Wahrscheinlichkeiten der Fehler vom Typ I und II entscheidend für die Qualität

- ✓ der Diagnose einer bestimmten Person,
- ✓ wie auch für den langfristigen Erfolg.

Im folgenden Beispiel vergleichen wir zwei Situationen, welche die gleichen langfristigen Fehler aufweisen, aber unterschiedliche Qualitäten von Einzelentscheidungen widerspiegeln.

Daraus ergibt sich die enorme Bedeutung der zugrunde liegenden a priori-Wahrscheinlichkeit einer Erkrankung.

Statistische Inferenz verstehen lernen: Der Fall der “a priori”-Wahrscheinlichkeit

Im folgenden Beispiel vergleichen wir zwei Situationen, welche die gleichen langfristigen Fehler aufweisen, aber unterschiedliche Qualitäten von Einzelentscheidungen widerspiegeln.

Daraus ergibt sich die enorme Bedeutung der zugrunde liegenden a priori-Wahrscheinlichkeit einer Erkrankung.

Die „a priori“ spielt eine wesentliche Rolle für die Inferenz,

- ob sie nun versteckt ist wie bei der klassischen Inferenz,
- oder ob sie explizit verwendet wird wie bei der Bayesianischen Inferenz.

Anstatt einen der Ansätze zu bevorzugen, sollten die Inferenzmethoden durch die Verwendung von bedingten Wahrscheinlichkeiten geklärt werden.

Statistische Inferenz verstehen lernen:

Die Fisher-Neyman-Kontroverse

Fisher (1935) verwendete die Nullhypothese H_0 und die direkte Wahrscheinlichkeit $P(x | H_0)$

- als **abstrakte Distanz zwischen den Daten x und H_0** .
- Bei Fisher ist das Signifikanzniveau vorgegeben, und **es fehlt jegliche Stichprobeninterpretation**, das heißt, es kann nicht als relative Häufigkeit verstanden werden.

Neyman (1928) bestand darauf, dass ein statistischer Test

- neben der Nullhypothese auch eine Alternative betrachten muss,
- und ein Test eine Entscheidung darstellt, bei der **zwei verschiedene Arten von Fehlern** auftreten, **die als relative Häufigkeiten auf lange Sicht zu verstehen sind**.

Heute verwenden wir eine Hybridversion, nämlich den abstrakten p -Wert nach Fisher, wir interpretieren ihn aber als relative Häufigkeit auf lange Sicht.

Fehler vom Typ I und II sind wenig hilfreich, weil sie sich auf ein künstliches Szenario der Testwiederholung beziehen statt auf die Hypothese selbst .

Statistische Inferenz verstehen lernen:

Risiko von Fehlentscheidungen mit einer Interpretation auf lange Sicht

- Fisher validiert eine singuläre Nullhypothese anhand empirischer „Beweise“.
- Neyman und Pearson (1928) dagegen entwickeln ihre Politik der wiederholten Tests im Kontext der entscheidungsorientierten industriellen Prozesssteuerung.

Im Neyman-Pearson-Kontext können wiederholte Entscheidungen getroffen werden, so dass Fehler durch langfristige relative Häufigkeiten interpretiert werden können.

Allerdings

- ✓ beschränkt sich diese Häufigkeitsdeutung auf die Situation **innerhalb** desselben Szenarios,
- ✓ und kann **nicht** zwischen Szenarien unterschiedlicher Qualität erstreckt werden.

Bei diesen Fehlern handelt es sich um bedingte Wahrscheinlichkeiten und keineswegs um absolute Wahrscheinlichkeiten.

Statistische Inferenz verstehen lernen: Analogie zwischen Medizin und statistischen Tests

Risiko von Einzelentscheidungen in der medizinischen Diagnose:

H_0 der Patient hat diese Krankheit *nicht* (No Ca \equiv Kein Carcinom)

H_1 der Patient hat diese Krankheit (Ca \equiv Carcinom).

Die Diagnose positiv (+) ist ein Indiz, dass die Person die Krankheit hat.

Die Diagnose negativ (–) ist ein Indiz für das Gegenteil.

Fehler vom Typ I, auch α -Fehler = $P(+ | H_0) = 1 - \text{Spezifität}$

Der Patient erhält die Diagnose (+), was interpretiert wird, als ob er diese Krankheit hätte, obwohl er eigentlich frei davon ist.

Fehler vom Typ II, auch β -Fehler = $P(- | H_1) = 1 - \text{Sensitivität}$

Der Patient erhält die Diagnose (–), was interpretiert wird, als ob er frei von dieser Krankheit wäre, obwohl er sie tatsächlich hat.

Statistische Inferenz verstehen lernen:

Analogie zwischen Medizin und statistischen Tests

In einer Klinik für Radiologie und in Vorsorgeuntersuchungen könnten wir Daten wie in den Folien weiter unten haben.

In der medizinischen Terminologie spricht man von

- ✓ **Spezifität** für die Wahrscheinlichkeit einer negativen Diagnose im Falle von H_0 (keine Krankheit),
- ✓ **Sensitivität** für die Wahrscheinlichkeit einer positiven Diagnose im Falle von H_1 (Krankheit).

Falsch-positive und falsch-negative Ergebnisse entsprechen Fehlern vom Typ I und II.

Es gibt aber eine viel relevantere Wahrscheinlichkeit:

Die Wahrscheinlichkeit, dass ein Patient bei einer positiven Diagnose (+) tatsächlich erkrankt ist, was einer isolierten Einzelfallentscheidung entspricht.

Statistische Inferenz verstehen lernen:

Analogie zwischen Medizin und statistischen Tests

Diese eminent wichtige Wahrscheinlichkeit für eine positive Diagnose wird als **positiver prädiktiver Wert (PPV)** bezeichnet.

Ähnlich verhält es sich mit dem **negativen prädiktiven Wert (NPV)** für eine negative Diagnose.

- Die Fehler vom Typ I und II beschreiben eher unbedeutende Aspekte der Qualität des diagnostischen Verfahrens.

Im folgenden Beispiel haben die klinische und die Screening-Situation die gleiche Sensitivität und Spezifität (gleiche Fehler vom Typ I und II).

Die Szenarien unterscheiden sich jedoch stark im prädiktiven Wert, entweder positiv oder negativ. PPV und NPV hängen vom jeweiligen Kontext ab:

- Eine positive Diagnose mag in der Klinik hilfreich sein, beim Screening ist sie nutzlos.

Fehlendes Bindeglied: Die Prävalenz und Bayes

Radiologische Klinik

	-	+	
H ₀ : No Ca	96	4	100
	Spezifität →	Falsch pos. →	
H ₁ : Ca	20	80	100
	Falsch neg. →	Sensitivität →	
	116	84	200

Screening

	-	+	
H ₀	95 232	3 968	99 200
	NPV ↑		
H ₁	160	640	800
		PPV ↑	
	95 392	4 608	100 000

Prävalenz

Sensibilität →

Spezifität →

Pos. prädikt. Wert (PPV) ↑

Neg. prädikt. Wert (NPV) ↑

50 %

80/100 = **80.0%**

96/100 = **96.0%**

80/84 = **95.2%**

96/116 = 82.8%

0.8 %

80/80 = **100.0%**

96/95 392 = **0.1%**

160/640 = **25.0%**

99 200/99 200 = **100.0%**

P (+ | Ca)

P (- | No Ca)

P (Ca | +)

P (No Ca | -)

Fehlendes Bindeglied: Die Prävalenz und Bayes

Radiologische Klinik

	-	+
H ₀ : No Ca	96 Spezifität →	4 Falsch pos. →
H ₁ : Ca	20 Falsch neg. →	80 Sensitivität →
	116	84

Screening

	-	+
	3 968	99 200
	4 608	100 000

Die Qualitätsindices haben in beiden Situationen denselben Wert:

- Sensibilität = 80%
- Spezifität = 96%

Aber die Qualität der Entscheidungen ist sehr verschieden:

- In der Klinik vielleicht eine Unterstützung.
- Im Screening vollkommen obsolet.

Die Diskrepanz ist umso größer, je kleiner die Prävalenz ist.

Prävalenz

Sensibilität →

Spezifität →

Pos. prädikt. Wert (PPV) ↑

Neg. prädikt. Wert (NPV) ↑

50 %

80/100 = 80.0%

96/100 = 96.0%

80/84 = 95.2%

96/116 = 82.8%

0.0 %

80/100 = 80.0%

96/100 = 96.0%

13.9%

99.8%

P (+ | Ca)

P (- | No Ca)

P (Ca | +)

P (No Ca | -)

Fehlendes Bindeglied: Die Prävalenz und Bayes

Radiologische Klinik

	-	+
H ₀ : No Ca	96 Spezifität →	4 Falsch pos. →
H ₁ : Ca	20 Falsch neg. →	80 Sensitivität →
	116	84

Screening

A posteriori Wahrscheinlichkeiten:
 haben ohne Bezug auf a priori Wahrscheinlichkeiten keinerlei Sinn,
 entbehren über ein sehr eng abgestecktes Szenario hinaus
 vollständig einer Häufigkeitsdeutung.

Überdies ist diese Häufigkeitsdeutung für das Individuum wertlos,
 sie dient nur als Referenzpunkt für das System
 (auch als Argumentationshilfe,
 um die diagnostische Prozedur zu "rechtfertigen").

Prävalenz

Sensibilität →

Spezifität →

Pos. prädikt. Wert (PPV) ↑

Neg. prädikt. Wert (NPV) ↑

	50 %	0.8 %
80/100 =	80.0%	80.0%
96/100 =	96.0%	96.0%
80/84 =	95.2%	13.9%
96/116 =	82.8%	99.8%

P (+ | Ca)

P (- | No Ca)

P (Ca | +)

P (No Ca | -)

3. Schlussfolgerungen

**Alle Wurzeln der Wahrscheinlichkeit nutzen,
um einen nachhaltigen Sinn zu schaffen
für statistische Schlussfolgerungen**

Schlussfolgerungen:

Klassische statistische Inferenz ignoriert die a priori-Wahrscheinlichkeiten

Die Kontroverse in den Grundlagen ist auf die a priori-Wahrscheinlichkeit zurückzuführen.

- Die Analogie zwischen statistischen Tests und medizinischer Diagnose hilft zu erkennen, dass bei klassischen Tests die a priori-Wahrscheinlichkeit der Nullhypothese fehlt.

Diese a priori-Wahrscheinlichkeit

- kann keine frequentistische Wahrscheinlichkeit sein,
- sondern muss einen qualitativen Grad der Überzeugung darstellen, also eine subjektivistische (epistemische) Wahrscheinlichkeit sein.

Ganz wie dies auch schon Brosseau mit seiner „glissement didactique“ als allgemeine Gefahr von didaktischen Übervereinfachungen dargestellt hat:

Die Vereinfachung der Komplexität der Inferenz führt zu einer Karikatur der Konzepte.

Schlussfolgerungen:

Das Problem der Reduktion der Komplexität

Die Reduzierung der Komplexität auf der Grundlage einer rein frequentistischen Wahrscheinlichkeit löst die konzeptionellen Probleme nicht.

Ebenso mag ein Bayesianischer Ansatz intuitiver sein und auf natürliche Weise zur statistischen Inferenz führen, aber auch hier gilt es, das gesamte Konzept der Wahrscheinlichkeit abzudecken.

Gestaltung eines umfassenderen Konzepts der statistischen Inferenz – und der Wahrscheinlichkeit:

Migon und Gamerman (1999) und Vancsó (2009) schlagen vor, die klassische und die Bayesianische Inferenz parallel zu unterrichten.

Schlussfolgerungen: Alle Deutungen von Wahrscheinlichkeit stützen und mit Fragen der statistischen Beurteilung verknüpfen

Zwei Aussagen von zukünftigen Lehrkräften von Vancsó könnten überzeugen:

„Ich habe das Konfidenzintervall erst verstanden, nachdem ich mich mit der Bayesianischen Region maximaler Dichte vertraut gemacht habe.“

„Ich mag Bayesianische Ideen sehr gerne ... weil ich dadurch gesehen habe ... warum Menschen unterschiedliche Meinungen haben ... weil sie eben unterschiedliche Vorverteilungen haben.“

Größere Komplexität ermöglicht ein tieferes Verständnis:

Die Herausforderung besteht darin, geeignete Lernwege zu entwickeln.

Schlussfolgerungen:

Weitere Herausforderungen – Der Fall der kleinen Wahrscheinlichkeiten

Es ist unmöglich, aus Daten zu einer Wahrscheinlichkeit von nur 10^{-4} zuverlässige Informationen zu gewinnen, so dass kleine Wahrscheinlichkeiten empirisch nicht erfassbar sind sondern nur durch Annahmen modelliert werden können.

- Dies führt zu einer eher qualitativen als zu einer frequentistischen Konnotation von Wahrscheinlichkeit.
- Die statistische Inferenz ist gespickt mit kleinen Wahrscheinlichkeiten.

Entweder wir verwenden Annahmen in Form von Modellen
oder wir verweisen auf den metaphorischen Charakter von Wahrscheinlichkeit.

Schlussfolgerungen:

Die Logik wiederholter Entscheidungen

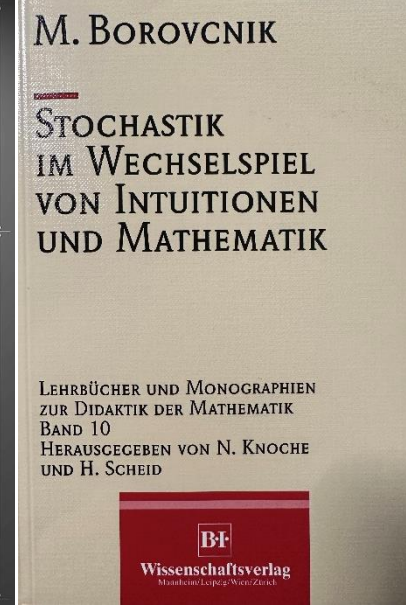
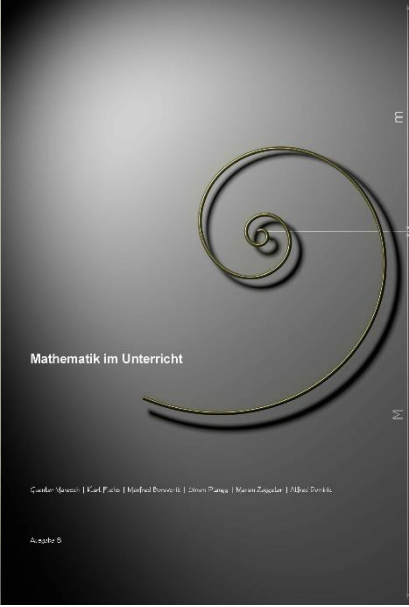
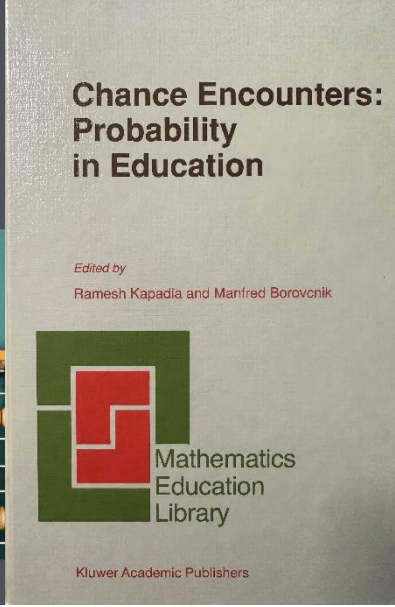
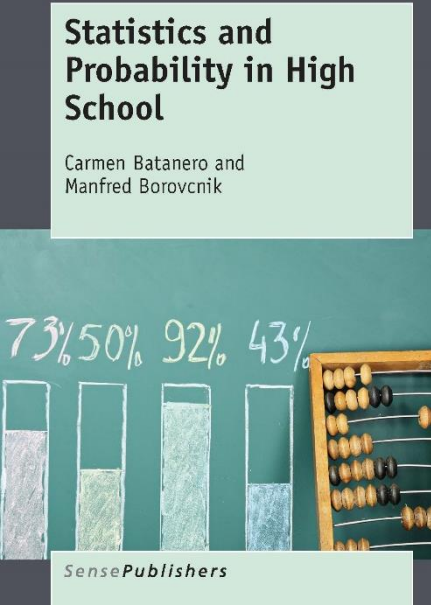
Eine optimale Entscheidung hängt davon ab, ob sie

- einmal (ein einziges Mal)
- oder wiederholt getroffen wird.

Im medizinischen Bereich unterscheidet sich die Entscheidung eines Staates oder einer Institution von der optimalen Entscheidung einer Privatperson aufgrund

- ✓ der unterschiedlichen Logik von wiederholten Entscheidungen im Vergleich zu Einzelentscheidungen,
- ✓ Unterschiede im Nutzen und die Interessen der Beteiligten auf verschiedenen Ebenen kommen noch verschärfend hinzu.

Wir befürworten eine pluralistische Perspektive auf den Begriff Wahrscheinlichkeit, die eine vergleichende statistische Inferenz (Barnett, 1982) beinhaltet und auf den Bildungsbereich übertragen wird.



Danke für Ihr Interesse!

Mehr finden Sie unter:

www.researchgate.net/profile/Manfred-Borovcnik

manfred.borovcnik@aau.at



Literatur

Allgemeine Einträge

- Barnett, V. (1982). *Comparative statistical inference* (2nd ed.). Wiley.
- Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Sense Publishers.
- Carranza, P., & Kuzniak, A. (2008). Duality of probability and statistics teaching in French education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table*. ICMI/IASE. www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf
- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum. *Technology Innovations in Statistics Education*, 1(1), 1–15. escholarship.org/uc/item/6hb3k0nz#page-1
- delMas, R. (2017). *A 21st century approach towards statistical inference—Evaluating the effects of teaching randomization methods on students' conceptual understanding* [Paper presentation]. 61st ISI World Statistics Congress, Marrakesh, Morocco.
- Diepgen, R. (1992). Objektivistische oder subjektivistische Statistik? Zur Überfälligkeit einer Grundsatzdiskussion. *Stochastik in der Schule*, 12(3), 48–54.
- Fejes-Tóth, P., Vancsó, Ö., & Borovcnik, M. (2022). Combinatorial thinking as key for introducing hypothesis testing—evaluation of the planned secondary-school reform in Hungary. In S.A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the gap: Empowering and educating today's learners in statistics (ICOTS 11)*. IASE.

Literatur

Allgemeine Einträge

- Finetti, B. de (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré*, 7(1), 1–68.
- Fisher, R. A. (1971). *The design of experiments*. Oliver & Boyd. (Original work published 1935)
- Hacking, I. (1965). *The logic of statistical inference*. Cambridge University Press.
- Hoegh, A. (2020). Why Bayesian ideas should be introduced in the statistics curricula and how to do so. *Journal of Statistics Education*, 28(3), 222–228. doi.org/10.1080/10691898.2020.1841591
- Hubbard, R., & Bayarri, M. J. (2003). Confusion over measures of evidence (p) versus errors (β) in classical statistical testing. *The American Statistician* 57(3), 171–182. doi.org/10.1198/0003130031856
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292.
- Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd English ed.; N. Morrison, Trans). (Original work published 1933). Chelsea.
- Migon, H. S., & Gamerman, D. (1999). *Statistical inference: An integrated approach*. Arnold.
- Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Part I and II. *Biometrika*, 20A, 175–240; 263–294.

Literatur

Allgemeine Einträge

- Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia & M. Borovcnik, (Eds.), *Chance encounters* (pp. 135–167). Kluwer. doi.org/10.1007/978-94-011-3532-0_5
- Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal of Mathematics Education* 4(3), 181–212.
doi.org/10.29333/iejme/242
- Varga, T. (1983). Statistics in the curriculum for everybody—How young children and how their teachers react. In D. R. Grey, P. Holmes, V. Barnett, & G. M Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (Vol. 1, pp. 71–80). Teaching Statistics Trust.
iase-web.org/Conference_Proceedings.php?p=ICOTS_1_1982
- Witmer, J., Short, T. H., Lindley, D. V. Freedman, D. A., & Scheaffer, R. L. (1997). Teacher's corner. Discussion of papers by D. A. Berry, J. Albert, & D. S. Moore. *The American Statistician*, 51(3), 262–274.
doi.org/10.2307/2684895

Literatur

Autor

- Borovcnik, M. (2021). The meaning of probability from a foundational perspective. *Revista Sergipana de Matemática e Educação Matemática* 6(3), 42-76.
- Borovcnik, M. (2021). Mutual Influence between different views of probability and statistical inference. *Revista Paradigma, Edición Especial – Educación Estadística*, 42(Extra 1), 221-256.
- Borovcnik, M. (2021). Corner pillars of probability literacy. *Proc. 63rd ISI World Statistics Congress*. ISI. www.isi-web.org/files/docs/papers-and-abstracts/154-day3-ips078-corner-pillars-of-probability.pdf
- Borovcnik, M. (2020). Contextos, analogías y tareas, que exponen el propósito de los conceptos clave de probabilidad. *Zetetiké* 28, 1-24. periodicos.sbu.unicamp.br/ojs/index.php/zetetike/issue/view/1648
- Borovcnik, M. (2019). Conceptos fundamentales y sus propiedades clave en probabilidad – Cómo identificarlos y proporcionar intuiciones que los sostengan. *REVEMAT* 14, *Edição Especial*, 1-25. funes.uniandes.edu.co/27163/1/Borovcnik2019Conceptos.pdf
- Borovcnik, M. (2019). Inferencia informal e inferencia “informal”. *Educação Matemática Pesquisa* 21(1), 433-460. doi.org/10.23925/1983-3156.2019v21i1p433-460
- Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk. *Educação Matemática Pesquisa*, 18(3), 1491-1516.
- Borovcnik, M. (2015). Risk and decision making: The “logic” of probability. *The Mathematics Enthusiast*, 12(1–3), 113–139. doi.org/10.54870/1551-3440.1339